

Nature and causes of trends in male diabetes prevalence, undiagnosed diabetes, and the socioeconomic status health gradient

James P. Smith*

Rand Corporation, 1776 Main Street, Santa Monica, CA 90407

Edited by Richard A. Easterlin, University of Southern California, Los Angeles, CA, and approved June 20, 2007 (received for review December 18, 2006)

This paper investigates levels in diabetes prevalence patterns across key socioeconomic status indicators and how they changed over time. The investigation spans both the conventional concept of diagnosed diabetes and a more comprehensive measure that includes those whose diabetes is undiagnosed. By doing so, I separate the distinct impact of covariates on trends over time in disease onset and the probability of disease diagnosis. The principal force leading to higher diabetes prevalence over time is excessive weight and obesity, which was only partially offset by improvements in the education of the population over time. Undiagnosed diabetes remains an important health problem, but much less so than 25 years ago. Although race and ethnic differentials in undiagnosed diabetes were eliminated over the last 25 years, the disparities became larger across other measures of disadvantage, such as education.

diagnosed

Diabetes is a serious illness dealing with the body's inability to produce (type 1) or regulate (type 2) insulin, which controls the level of glucose in the blood. Diabetes prevalence rises rapidly with age, is believed to be increasing rapidly over time (1), and is high among Americans (2). The consequences of diabetes can be quite severe, including heart and kidney disease, poor circulation occasionally resulting in amputation of limbs, vision problems with blindness a possibility, a diminished quality of life, and premature death (1). Although there is evidence that the mortality and cardiovascular disease (CVD) risk associated with diabetes may have declined over time (3), in part due to better health care (4), diabetes remains a serious disease.

Research has indicated that the incidence and prevalence of diabetes is more common among those at the bottom of several prominent socioeconomic status (SES) markers, such as education and income (5), as well as for America's principal ethnic and racial minorities: African-Americans and Hispanics.

In this paper, I investigate the patterns of diabetes prevalence overall and across key SES indicators and how those patterns have changed over time. My investigation spans both the conventional concept of diagnosed diabetes and a more comprehensive measure that includes those whose diabetes is as yet undiagnosed. By doing so, I can separate the distinct impact of covariates on disease onset and the probability of disease diagnosis. Special emphasis is given to SES correlates of undiagnosed diabetes and how these changed as the fraction of Americans with undiagnosed diabetes plummeted over the last 25 years.

Data, Measures, and Methods

This research uses various waves of the National Health and Nutrition Examination Surveys (NHANES): NHANES IV (1999–2002), NHANES III (1988–1994), and NHANES II (1976–1980). NHANES contain data obtained through personal interviews, physical exams, and laboratory tests. All data are available for adults ages 25–70, the age span in this analysis. Details of the specific survey and sampling procedures used can be found in the

references cited for each NHANES (6–9). NHANES II oversampled low-income households whereas the latter two NHANES oversampled African-Americans and Hispanics. All tabular data are weighted.

In all waves, information is available on self-reported prevalence of many illnesses, including diabetes of the form “Did a doctor ever tell you that you had diabetes. . .” NHANES does not allow a distinction between type 1 (insulin deficient) and type 2 (insulin resistant) diabetes, although well over 90% of diabetics in an adult population are type 2. Individual attributes including age, gender, race, marital status, family income, education, and parental diabetes are obtained from individual interviews. Unfortunately, gestational diabetes, a significant component of diabetes among women, is neither consistently included nor excluded in NHANES over time, and thus this research will focus on men only (10).

The key advantage of NHANES is that all waves contain data obtained both through physical tests and laboratory exams (blood, urine, and swabs). Particularly relevant for research with a focus on diabetes, physical tests were performed on height and weight so that body mass index (BMI) could be computed and with it an objective indicator of whether or not the respondent is obese ($\text{BMI} \geq 30$) or overweight ($\text{BMI} \geq 25$ and <30). I separate the obese population into three subgroups: class 1 (≥ 30 and <35), class 2 (≥ 35 and <40), and morbid (≥ 40) (11).

The laboratory examination for diabetes in NHANES III and IV is a glycosylated hemoglobin (HbA1c) test, a measure of the percent of hemoglobin molecules bound to glucose. Although not usually a screener for diabetes, HbA1c is highly correlated with fasting plasma glucose levels (12, 13). Although there is no strict diagnosis threshold value, I follow the standard convention by using values $\geq 6.5\%$ as indicating clinical diabetes. My principal results are not sensitive to the specific threshold chosen.

NHANES II relied instead on fasting glucose. For NHANES II, I defined clinical diabetes by using a classification of the Oral Glucose Challenge/Tolerance Test (OGTT) developed by NHANES. In part, the OGTT involves measurement of plasma glucose concentrations in the fasting state and cut by whether the fasting level is ≥ 140 mg/dl. Laboratory tests in NHANES II were not given to all sample participants, but only those randomly selected for blood tests. Many self-reported diagnosed diabetics were not tested, and other respondents had to be excluded from the laboratory analyses because their tests results were not useable. Because many self-reported diabetes were not tested, one cannot in NHANES II do analysis on prevalence of clinical tested diabetes only. In addition, sample sizes for analyses that rely on the labo-

Author contributions: J.P.S. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; NHANES, National Health and Nutrition Examination Surveys; SES, socioeconomic status.

*E-mail: smith@rand.org.

© 2007 by The National Academy of Sciences of the USA

Table 1. Male diabetes prevalence rates and diabetes risk factors by race and ethnicity: 25–70 years old

Prevalence and risk factors	All races				White non-Hispanics				African-Americans				Hispanics			
	1999–2002	1988–1994	1976–1980		1999–2002	1988–1994	1976–1980		1999–2002	1988–1994	1976–1980		1999–2002	1988–1994	1976–1980	
Prevalence																
Diagnosed, %	7.0	4.6	3.1		6.3	4.7	3.0		8.4	5.7	5.0		8.5	3.9	1.4	
Clinical, %	7.1	5.2	NA		6.4	4.8	NA		8.5	8.8	NA		7.6	5.3	NA	
Total prevalence, %	8.9	6.8	6.0		8.0	6.3	5.6		11.1	10.3	8.4		10.8	7.0	4.0	
Undiagnosed, %	21.6	32.5	48.2		21.2	26.4	46.0		24.3	45.1	40.3		21.4	44.0	65.4	
Risk factors																
Obese (clinical), %	28.2	21.0	13.1		28.8	21.3	12.8		29.2	21.5	17.2		24.3	23.0	15.1	
Height (clinical), in	69.4	69.3	69.0		69.9	69.6	69.2		69.7	69.5	69.0		67.1	66.9	67.0	
Age, years	44.6	43.1	44.4		45.7	43.8	44.6		43.9	41.8	43.5		40.8	40.1	42.0	
Low ed, %	20.9	23.1	33.2		12.8	18.0	29.4		36.4	32.3	53.0		46.7	56.5	69.8	
Middle ed, %	24.9	31.1	31.3		26.4	32.0	33.1		22.7	36.9	24.9		19.2	21.6	17.3	
High ed, %	55.1	45.7	35.4		60.8	50.0	37.5		40.9	30.7	22.1		33.9	21.9	12.9	

Source: NHANES II (1976–1980), III (1988–1994), and IV (1999–2002). All data are weighted. Race and Hispanic ethnicity is ascertained from a self-report of respondents. Diagnosed prevalence, whether a doctor told the respondent that they were diabetic; clinical prevalence, HbA1c level $\geq 6.5\%$; total prevalence, either self-report or clinical; undiagnosed, the fraction of total prevalence that is not diagnosed; obese, BMI ≥ 30 ; low ed, those who did not graduate high school; middle ed, high school graduates; high ed, those who went beyond high school.

ratory test results are much smaller in NHANES II, limiting analysis by using laboratory results.

This series of NHANES allows one to monitor 25-year trends in diagnosed and total diabetes (either diagnosed or undiagnosed). The total measure codes a person as a diabetic if either he self-reported that a doctor told him he was diabetic or he had a clinical value above the diagnostic threshold. Diagnosed diabetes would not track trends in disease prevalence because as documented below almost half of diabetics were undiagnosed in the late 1970s, a fraction that has declined sharply over time. However, a clinical measure only would also not be correct. It would exclude the nontrivial fraction (one-quarter or more) of diabetics who have their disease currently below the diagnostic threshold due to good disease management.

Some self-reported diabetes may be in error if a respondent incorrectly answered the question. To examine test–retest error in self-reports, I used the Health and Retirement Survey (HRS) (<http://hrsonline.isr.umich.edu>). In a sample of those in the original HRS (ages 51–61 years old in the baseline sample), 9% of those who previously self-reported they were diabetic denied that report 2 years later. That amount of test–retest error would have a very small effect on total prevalence rates reported in this paper. For these reasons, the total diabetes prevalence measure is seen as most accurately measuring the actual prevalence of disease in the population.

The combination of biological measures and self-reports of disease prevalence permits tracking differential trends in undiagnosed diabetes. Age-specific diabetes prevalence is modeled as a function of indicators for race and ethnicity, a quadratic in age, ever and current smoking, vigorous physical activity, overweight, the three stages of obesity class, height (in inches), and parental prevalence of diabetes (either the father or mother was a diabetic). Probit models were used, but the main conclusions do not depend on the specific statistical model chosen.

All versions of NHANES collect several health-related behaviors thought to be significant risk factors for diabetes. These include whether the respondent ever was or is a current smoker and the amount of vigorous physical exercise in which one normally engages. The definition of exercise varies by wave. In NHANES IV, respondents were asked: “Over the last 30 days, did you do any vigorous activities for at least 10 min that caused heavy sweating or large increases in breathing or heart rates? Some examples are running, lap swimming, aerobics classes or fast bicycling.” In NHANES II, respondents were asked: “In the things you do for recreation, for example, sports, hiking, dancing and so forth, do you get much exercise, moderate exercise, or no exercise”; only “much” is counted as vigorous. In NHANES III, one is asked how often over the last month you did a set of physical activities. An intensity-rated scale is given to each activity. I examined the intensity scales of activities counted as vigorous in NHANES III, and only activities in NHANES III receiving that score or above are counted as vigorous exercise. Although these revisions create imprecision in the amount of real change in vigorous exercise over time, these exercise variables provide the same threshold at a point in time so that their impact on the diabetes gradient can be ascertained.

My main SES measure is years of education, dividing men into three education groups: less than, equal to, or more than a high school education. For some analyses using NHANES II and III, I can separate college graduates from those with 13–15 years of schooling, a division not possible in NHANES IV. Because NHANES places annual family income categorized into relatively few income categories, NHANES is not the data from which to conduct a detailed investigation of the role of income. I divide total family income into three approximately equally sized groups or terciles. That goal produced income tercile cutoffs points of \$35,000 and \$65,000 in NHANES IV, \$25,000 and \$50,000 in NHANES III,

and \$10,000 and \$20,000 in NHANES II. Race and Hispanic ethnicity is ascertained from a self-report of respondents.

Measures of obesity, overweight, and height used are based on physical exams of respondents and are free of the well known measurement errors associated with self-reports. For each NHANES, comparable models are estimated both for self-reports of diabetes prevalence and the more comprehensive total prevalence measure. I also estimate models of the probability of having undiagnosed diabetes, conditional on being a diabetic.

Results

Table 1 documents trends in alternative measures of diabetes prevalence for all men 25–70 years old and for major ethnic and racial groups separately. Examining diagnosed diabetes, the increase in prevalence is dramatic. The fraction of men diagnosed as diabetics has more than doubled from 3.1% to 7.1%. In the most recent NHANES, diagnosed diabetes is approximately one-third more prevalent among African-Americans and Hispanics than among non-Hispanic White men.

Examining total prevalence obtained by combining clinically evaluated and diagnosed diabetes, male diabetes is higher than diagnosed diabetes. This makes this serious disease more common with approximately 1 in every 11 men afflicted between 1999 and 2002. At the same time, secular trends, although still real and significant, are far less dramatic. Among men of all races, diabetes prevalence rose over this period from 6% to 9%, a 50% increase compared with more than a doubling when diagnosed diabetes was used. The reason for the discrepancy results from a large fall in the fraction of undiagnosed. Over these 25 years, rates of male undiagnosed diabetes fell sharply from almost half to a little more than one in five. Especially during the 1990s, these declines in undiagnosed diabetes were large among Hispanic and African-American men.

Table 1 also documents levels and trends in some prominent diabetes risk factors. Male heights increased by less than one-half inch while mean age was also relatively constant over time. The main exception to this stagnant portrait is that male obesity more than doubled, growing from 13% to 28% (13). Secular trends in education are strongly positive as the fraction of men without a high school diploma fell from one-third to one-fifth, whereas the percentage who went beyond high school increased from 35% to 55%. The steady secular advance in education accomplishments of all male ethnic groups over time was only partially offset by the increased weight given to Latinos as their immigration increased dramatically over this period.

This research attempts to understand reasons for the education diabetes gradient and how it evolved over time (5). To set background, Table 2 provides some key outcomes stratified by education. Especially in recent years, there exists a pronounced negative gradient of diagnosed diabetes with schooling. In NHANES IV, 6% of men who went beyond high school had diagnosed diabetes compared with 9.8% of men who did not obtain a high school degree. In contrast, the education gradient in diagnosed diabetes in NHANES II is much more muted. In fact, controlling for age, there is no education gradient in diagnosed diabetes in NHANES II. For men ages 55–70, the fraction with diagnosed diabetes was 6.4%, 6%, and 7.2%, among those with less than a high school degree, only a high school degree, and more than a high school degree, respectively. In each NHANES, education gradients are much sharper in the more comprehensive and presumably more accurate total prevalence measure.

The other rows in Table 2 suggest possible reasons for the education gradient in diabetes prevalence and how it evolved over time. Those in the lowest education group are more likely to be Latino or African-American, less likely to engage in vigorous physical exercise, much more likely to smoke, and more likely to be obese. Although the percentage with a parent who was a diagnosed

Table 2. Patterns of male diabetes prevalence and diabetes risk factors by education group and NHANES survey year

Prevalence and risk factors	Education level					
	Low	Middle	High	All	Ed 13–15	Ed 16+
NHANES II: Ages 25–70						
Prevalence						
Diagnosed prevalence, %	3.9	3.4	2.0	3.1	2.0	1.9
Total prevalence, %	8.1	6.6	3.1	5.8	1.9	3.9
Undiagnosed, %	48.2	43.7	49.6	47.9	51.2	49.1
Risk factors						
Hispanic, %	9.9	2.6	2.5	4.7	2.3	2.6
Black, %	14.6	7.3	5.7	9.3	8.3	3.8
Now smoker, %	49.1	44.8	35.2	42.8	42.8	29.7
Vig-exercise, %	20.9	23.4	23.9	22.8	22.6	24.9
Obese, %	15.4	15.5	9.1	13.1	13.1	7.1
Parent diabetic, %	17.8	16.8	15.2	16.5	15.7	15.1
Age, years	49.7	42.8	41.9	44.4	41.1	40.3
NHANES III: Men ages 25–70						
Prevalence						
Diagnosed prevalence, %	7.2	4.4	3.4	4.6	4.4	2.6
Total prevalence, %	12.1	6.4	4.3	6.8	5.8	3.3
Undiagnosed, %	40.5	30.7	22.9	32.5	24.6	20.7
Risk factors						
Hispanic, %	23.0	6.4	4.4	9.5	6.8	3.3
Black, %	13.6	11.6	6.6	9.9	9.6	4.3
Now smoker, %	44.9	42.8	21.0	33.7	27.6	15.9
Vigorous-ex, %	20.5	29.3	50.8	37.0	43.3	56.4
Obese, %	24.7	22.2	18.6	21.1	21.7	16.1
Parent diabetic, %	21.9	20.4	18.6	19.9	19.7	17.7
Age, years	46.3	42.2	42.1	43.1	40.7	43.1
NHANES IV: Men ages 25–70						
Prevalence						
Diagnosed prevalence, %	9.8	7.0	6.0	7.0		
Total prevalence, %	14.2	8.7	7.2	8.9		
Undiagnosed, %	31.1	19.3	16.4	21.8		
Risk factors						
Hispanic, %	32.4	10.7	7.2	13.9		
Black, %	17.6	8.8	8.5	9.6		
Now smoker, %	43.2	36.6	18.9	28.2		
Vig-exercise, %	23.5	33.5	51.4	41.4		
Obese, %	28.3	31.1	26.9	28.2		
Parent diabetic, %	26.1	25.4	23.4	24.4		
Age, years	45.3	43.6	44.9	44.6		

Now smoker, currently a smoker; Vig-exercise, vigorous exercise as defined in the text; Obese, BMI ≥ 30 ; parent a diabetic, one of the respondent's parents was diabetic. All other variables are defined in Table 1. All data are weighted.

diabetic does decline with education, the strength of this relationship is weak.

I turn next to the risk factors associated with schooling that have changed the most over time. First on that list would be the increasing fraction of Latinos in the lower education groups. In NHANES IV, among men who were not high school graduates, one-third were Hispanics; in NHANES II, the comparable fraction was 10%.

The second important factor is age. There was a >9-year age difference between college graduates and non-high-school grads in NHANES II and a <2-year difference now. The negative education smoking gradient also became much larger over time. Although rates of obesity have risen over the last 25 years, there is no steepening of that education gradient. Education gradients in parental diabetes also seem not to have been altered over time.

Table 3 lists by education, income, and ethnicity rates of undiagnosed diabetes in each NHANES. Especially in NHANES III and IV, there is a sharp negative gradient in undiagnosed diabetes across education and income groups (14). Using NHANES III to illustrate, 41% of diabetics with less than a high school degree are unaware of their condition. The comparable fraction for those

Table 3. Percentage of men 25–70 with diabetes who have undiagnosed diabetes in NHANES II, III, and IV

Variable	NHANES II (1976–1980)	NHANES III (1988–1994)	NHANES IV (1999–2002)
Education, %			
Ed 0–11	48.2	40.5	31.1
Ed 12	43.7	30.7	19.3
Ed > 12	49.6	22.9	16.4
Income, %			
Lowest	55.1	37.8	27.5
Middle	44.3	26.3	13.8
Highest	45.5	25.7	19.4
Ethnicity, %			
White non-Hispanic	46.0	26.4	21.2
Hispanic	65.4	44.0	21.4
African-American	41.7	45.1	24.3
All	48.2	32.5	21.6

Respondents are divided into three equal income terciles in each NHANES. All other variables are defined in Tables 1 and 2. All data are weighted.

beyond high school is only 23%. There was a large decline in the percentage of diabetes undiagnosed falling from almost half to one-fifth. These improvements in eliminating undiagnosed diabetes appear to be larger among the more educated and to a lesser extent those with the most income. In contrast, the huge differentials in undiagnosed diabetes across racial and ethnic groups have for all practical purposes been eliminated.

Model Estimates

In this section, I discuss probit models of determinants of diagnosed and total diabetes prevalence among men for all three NHANES waves. The purpose of these models is twofold. The first is to identify the most important factors leading to higher probability of diabetes, and the second is to ascertain how the influence of these factors may have changed over time in their relative importance.

Table 4 contains estimates of two diabetes prevalence models in three NHANES waves. By including both diagnosed and undiagnosed diabetes, the total prevalence model in the final two columns comes closer to estimating the relationship of covariates to the actual presence of diabetes among men aged 25–70. Estimates in the first two columns tell us about the relation of covariates to diagnosed diabetes only. The differences between effects in diagnosed and total prevalence models are indicative of the impact of covariates on the probability of undiagnosed diabetes. Table 4 lists estimated partial derivatives alongside the associated “z” statistic for the estimated effect being different from zero. Robust standard errors are used.

First, examine estimated effects for the total prevalence. Even after controlling for personal attributes, diabetes is significantly higher among both Latino and African-American men, but the estimated disparity is smaller in the most recent NHANES, especially for African-Americans. Diabetes prevalence increases with age, albeit at a decreasing rate. The probability of being a diabetic is much higher if a parent was diabetic. The extent to which this generational transmission reflects common genetic influences or a shared family social and environmental background is not knowable from these estimates alone.

Being either overweight or obese raises the odds that one is diabetic with the estimated obesity impact being much larger than overweight. These estimated impacts of obesity on prevalence are one reason why obesity is associated with excess mortality (15). The size of the estimated effects increases across the three stages of obesity. The effects of excessive weight are much stronger in the combined total prevalence estimates, an issue to which I return below. Engaging in vigorous exercise is negatively associated with diabetes, but its impact does not depend on the definition of diabetes prevalence. In none of the

Table 4. Probit models for male prevalence of diagnosed and total diabetes prevalence: Ages 25–70

Variable	Diagnosed		Total prevalence	
	dF/dx	z	dF/dx	z
NHANES IV				
Hispanic	0.015	1.37	0.019	1.52
Black	0.017	1.57	0.028	2.19
Age	0.012	3.78	0.016	4.12
Age2	−0.001	2.63	−0.0001	2.77
Ed mid	−0.004	0.33	−0.016	1.26
Ed high	−0.005	0.50	−0.020	1.69
Income mid	−0.010	1.10	−0.016	1.47
Income high	−0.038	3.54	−0.047	3.73
Ever smoked	0.004	0.50	0.011	1.07
Current smoker	−0.014	1.44	−0.016	1.37
Vig exercise	−0.013	1.52	−0.016	1.58
Overweight*	0.016	1.47	0.039	2.96
Obesity 1*	0.024	1.75	0.069	3.94
Obesity 2*	0.096	4.45	0.198	6.88
Obesity 3*	0.152	4.73	0.288	7.02
Height*	−0.002	1.38	−0.002	1.43
Parent diabetic	0.071	7.36	0.080	7.18
n	3,106	3,106	3,109	3,109
NHANES III				
Hispanic	0.015	2.35	0.026	3.06
Black	0.020	3.09	0.050	5.70
Age	0.007	3.74	0.012	4.95
Age2	−0.000	2.42	−0.0001	3.36
Ed mid	−0.005	0.81	−0.011	1.47
Ed 13–15	−0.001	0.23	−0.012	1.37
Ed 16+	−0.006	0.79	−0.025	2.45
Income mid	−0.006	1.04	−0.003	0.39
Income high	−0.013	1.98	−0.016	1.69
Ever smoked	0.016	3.18	0.018	2.58
Current smoker	−0.019	3.64	−0.021	2.98
Vig exercise	−0.011	1.79	−0.017	2.23
Overweight*	0.015	2.48	0.031	3.81
Obesity 1*	0.034	4.12	0.081	7.00
Obesity 2*	0.066	4.19	0.166	7.25
Obesity 3*	0.139	5.20	0.279	7.63
Height*	−0.001	0.65	0.001	0.50
Parent diabetic	0.057	8.91	0.078	9.62
n	5,419	5,419	5,426	5,426
NHANES II				
Hispanic	−0.010	0.97	0.012	0.41
Black	0.027	3.66	0.046	1.99
Age	0.003	2.07	0.000	0.10
Age2	−0.000	0.87	0.000	0.79
Ed mid	0.010	1.95	0.023	1.60
Ed 13–15	0.005	0.63	−0.018	0.84
Ed 16+	0.008	1.05	0.011	0.51
Income mid	−0.010	2.14	−0.021	1.53
Income high	−0.011	1.98	−0.032	1.98
Ever smoked	0.001	0.14	−0.032	2.02
Current smoker	−0.009	1.89	0.005	0.34
Vig exercise	−0.013	2.46	−0.004	0.30
Overweight*	−0.000	0.08	0.030	2.31
Obesity 1*	0.007	0.95	0.030	1.31
Obesity 2*	0.095	4.00	0.028	3.03
Obesity 3*	0.157	3.18	0.664	3.58
Height*	0.001	1.11	−0.004	1.99
Parent diabetic	0.040	6.34	0.034	2.21
n	5,708	5,708	1,562	1,562

All models estimate robust standard errors. dF/dx are the estimated derivatives, and z are the test statistics for differences from 0. All variables are defined in the text and in Tables 1 and 2.

*Variables are objectively measured during physical exams.

models does height have any systematic association with being diabetic. Only in NHANES III is the positive estimated impact of ever smoking statistically significant, but this effect is negated by past smoking cessation.

Table 5. Estimated probit probability for both sexes of being an undiagnosed diabetic given that one is a diabetic

Variable	1999–2002		1988–1994	
	dF/dx	z	dF/dx	z
Hispanic	−0.006	0.16	0.025	0.63
Black	0.040	0.98	0.105	2.89
Female	−0.048	0.81	−0.108	1.93
Age	0.008	0.59	0.030	2.86
Age ²	−0.000	0.59	−0.000	2.90
Married	−0.068	1.34	−0.052	1.09
Married female	0.018	0.28	0.019	0.32
Ed mid	−0.053	1.33	0.029	0.87
Ed high	−0.084	2.36	0.000	0.02
Income mid	0.026	0.65	0.063	1.75
Income high	0.092	1.62	0.013	0.24
Ever smoked	0.038	1.06	−0.019	0.59
Current smoker	−0.015	0.35	0.048	1.30
Vig exercise	−0.023	0.57	0.078	1.70
Overweight*	0.102	1.69	0.037	0.85
Obesity 1*	0.133	2.05	0.152	3.24
Obesity 2	0.179	2.42	0.218	3.81
Obesity 3	0.153	1.95	0.175	2.80
BMI*–BMI (self)	−0.013	0.86	−0.003	0.25
BMI*–BMI (self) × obese	0.022	1.30	0.003	0.72
Height*	0.001	0.23	0.008	1.48
Parent diabetic	−0.068	2.24	−0.138	5.26
Have health insurance	0.030	0.75	−0.027	0.63
Last saw doctor 1–2 yrs	0.446	4.54	0.470	7.12
Last saw doctor 3 or more yrs	0.435	4.47	0.382	6.09
n	746	746	1,289	1,289

All models estimate robust standard errors. Indicator variables are included for whether respondent had health insurance and when they last saw a doctor with less than one year the left out category. All other variables are defined in Tables 1 and 2.

*Measured during physical exams.

Next, examine the difference between diagnosed and total prevalence models, which are indicative of the probability of being undiagnosed. The most systematic pattern is that effects of excessive weight are much larger in total prevalence models, indicating that obesity is strongly negatively correlated with the probability of being diagnosed (14). These results also suggest that race was associated with not being diagnosed but that its effect was eliminated over time.

Besides excessive weight, variables whose estimated impact differs the most between diagnosed and total prevalence models are the two SES markers: education and income. For both, I estimate a larger negative impact on prevalence when the most inclusive definition of prevalence is used that includes undiagnosed diabetes. For education, there exists no statistically significant association with diagnosed diabetes. In contrast, the estimated effects of education on prevalence are large and statistically significant when undiagnosed diabetes is included in prevalence. Especially for the middle-income group, a similar difference is found for the relation with income between the self-report and inclusive measure of prevalence.

The relation of these variables to undiagnosed diabetes is more directly captured in Table 5, which represents the probability of being an undiagnosed diabetic conditional on being a diabetic. Because this conditioning lowers sample sizes considerably, these models were estimated combining men and women. Tests for differences by gender did not indicate any significant differences between the male and female samples outside of an intercept shift. Even after combining the sexes, sample sizes in NHANES II were too small for any meaningful analysis, so these models were only estimated for NHANES III and NHANES IV.

A few additional variables were added to these models, including two meant to capture the extent of contact with the medical system.

These include whether one has health insurance and the last time one saw a doctor: 1–3 years and >3 years, with 1 year or less the omitted category. To capture the possible impact of misplaced self-perceptions, the difference between objective and self-reports of obesity also is included. This variable is interacted with an indicator variable that one is clinically obese. Finally, an indicator for marriage was added, an effect that is allowed to differ by gender.

The covariates that are not related to conditional nondiagnosis are smoking, exercise, Hispanic ethnicity, marriage, mistaken BMI perceptions, and height. In NHANES III, female diabetics were more likely to be diagnosed, and the probability of not being diagnosed increased with age. Although still present, both patterns were not statistically significant in NHANES IV.

Because doctors have standard checklists to query patients that include familial disease histories, it is unsurprising that having a diabetic parent reduces the likelihood that diabetes is not diagnosed. Parental diabetes is the best predictor for diabetes prevalence, and it is reassuring that it is taken into account in detection. But the second best predictor of undiagnosed diabetes is obesity, and it is not sufficiently taken into account. In all three stages of obesity and in both NHANES waves, the obese are more likely to be undiagnosed. Why this would be so is a bit of a mystery. One possibility is that the evidence relating obesity to diabetes is more recent. There may be (unnecessary) long lags in implementation into normal medical practice of using obesity as a signal of a potential problem.

Duration of time since last contact with a physician is positively related to being undiagnosed, although interpreting this effect is problematic because diagnosis induces additional physician visits. Health insurance is not related to the probability of diagnosis.

These results suggest a declining significance of race in being diagnosed. Race and ethnicity have been highlighted in National Institutes of Health (NIH) campaigns to reduce health disparities in health outcomes including disease detection and, based on these results, with some success. In NHANES III, African-Americans were more likely to not have been diagnosed, a result that is statistically significant. By NHANES IV, this disparity had disappeared. As race and ethnic disparities in diagnosis were eliminated, disparities by education appeared. In NHANES IV, diabetics in the highest education group are more likely to be diagnosed, a statistically significant difference not present in NHANES III. Health disparities appear in many ways, with race and ethnic differences more easy to monitor. Disparities across markers such as education level are more difficult to access and perhaps easier to ignore or dismiss as an underachieving patient problem. They are no less real. In neither of the two NHANES was income a marker for undiagnosed disease.

Explaining Trends in Diabetes Prevalence

In this section, my goal is to isolate factors most responsible for increasing diabetes prevalence over time (16). Let $P(A)$ and $P(B)$ be the (predicted) diabetes prevalence rates in years A and B and let $P(A)^{-j}$ and $P(B)^{-j}$ be the predicted prevalence in years A and B for a “counterfactual” situation that nobody suffers from risk factor j . $P(A) - P(A)^{-j}$ can be interpreted as the diabetes rate in year A due to that risk factor and similarly for year B .

The difference in diabetes prevalence in the 2 years is

$$P(B) - P(A) = [P(B)^{-j} - P(A)^{-j}] + [P(B) - P(B)^{-j}] - [P(A) - P(A)^{-j}].$$

The first term is the difference between diabetes prevalence in the two years not due to the risk factor j , whereas the sum of the second and third term is the part due to j . The latter two terms can each be separated in a “prevalence” effect (the percentage with the risk factor) and an “impact” effect (the impact of the risk factor on diabetes).

$$P(A) - P(A)^{-j} = 1/N_A \sum_{i \in A} \{g(x_i, b_A) - g(x_i^{-j}, b_A)\}$$

$$= \left[\sum_{i \in A} x_{ij} / N_A \right] \left[\sum_{i \in A, x_{ij}=1} \Delta g(x_i, b_A) / \sum_{i \in A} x_{ij} \right].$$

The first term is the fraction in year A that suffers from factor j . Because $\Delta g(x_i, b_A)$ is the marginal effect (“partial derivative”) for a dummy variable, the difference if it is set to 1 or 0, the second term is the average marginal effect for those who have the risk factor. The same decomposition can be used for all variables, allowing one to compare the importance of each to diabetes prevalence in each year and then to differences between the years.

Table 6 presents my accounting for the increase in total diabetes prevalence between the three NHANES waves. Some attributes either did not change over time (age and height) or had relatively small estimated effects on diabetes prevalence (smoking and exercise) and are excluded from this accounting. Those that remain include demographic factors (race and ethnicity), SES variables (income and education), and high levels of BMI.

Demographic forces had a relatively small overall impact. The small increase in prevalence predicted by the increasing numbers of Hispanics was partly offset by the diminished importance of race as a predictor of diabetes. Combined race and ethnicity predicts that diabetes prevalence would have risen only by 0.15 percentage points.

Increasing numbers of men who had a parent who was a diabetic and the high impact of parental diabetic inheritance combined to predict that male diabetes prevalence would have increased by 1.39 percentage points between the three NHANES data sets. Because parental diabetes only captures diagnosed diabetes among parents, some of the increase in parental prevalence reflects the reduction in undiagnosed diabetes among parents, and some reflects the growing actual prevalence among parents as diabetes prevalence grows over time.

Growth in excessive BMI was the most important factor leading to rising levels of diabetes over time. Being overweight was not critical (except by making one more likely to be obese in the future), but all three stages of obesity are important. The three stages combined predict an increase of diabetes of 2.02 percentage points, adding in the small contribution of overweight implies that excessive BMI leads to a 2.15 percentage point rise in male diabetes prevalence. Several papers have argued that the recent growth in obesity can at least partially be explained by declines in the relative price of food, reinforced by steep declines in the relative price of foods dense with calories (17).

The principal factor operating in the opposite direction was the improving levels of SES and most importantly higher levels of education. SES-related factors predict a decline in diabetes prevalence of 1.21 percentage points.

Total diabetes prevalence increased by 3 percentage points between the late 1970s and the beginning of this century. Three factors loom largest in explaining this increase: increasing obesity, followed by a rising fraction with a diabetic parent, with a quite small effect due to changing racial and ethnic demographics. The key offsetting force was the improving levels of SES. Combined, all these factors predict that diabetes prevalence would increase by 2.48 percentage points, 83% of the increase in male diabetes prevalence that actually took place.

If the objective instead was to explain the larger secular increase in diagnosed prevalence, then the large decline in undiagnosed diabetes would be added toward the top of the list of factors accounting for trends. Twenty-five percent of the increase in diagnosed diabetes since the late 1970s actually represents improved detection.

Table 6. Contribution of factors in explaining time series increase from 1976–1980 to 1992–2002 in male total prevalence of diabetes

Variable	Change, percentage points
Demographic variables	
Hispanic	0.19
Black	−0.04
Total demographics	0.15
Parent diabetic	1.39
Obesity variables	
Overweight	0.13
Obesity 1	0.72
Obesity 2	0.76
Obesity 3	0.54
Total obesity	2.02
Total obesity and overweight	2.15
SES variables	
Ed mid	−1.02
Ed-high	−0.95
Income mid	1.20
Income high	−0.44
Total SES	−1.21
All factors	2.48

Explaining the Education Diabetes Gradient

I turn now to isolating reasons that create the education health gradient in diabetes and why it changed over time. Although Kanjilal *et al.* (5) show similar trends over time in diabetes prevalence by education to those documented here, they do not model determinants of prevalence and only speculate about what may have caused its changing structure by education. To do so, I re-estimated total prevalence models in Table 5, first only controlling for education, the unadjusted education gradient. I then added variables in the following order, always maintaining the previous variables in the model: (i) age quadratic, race and ethnicity, (ii) smoking and exercise, (iii) parental diabetes, (iv) excessive weight, and (v) income groups. Variable i represents the demographically adjusted education gradient without controlling for any behavioral factors related to schooling and variable iv the adjusted gradient with only a single SES marker, schooling, without trying to parcel out the any distinct effects of schooling and incomes. This is my preferred model for understanding the nature of the schooling gradient with diabetes.

Estimates are provided in Table 7 for all six rows for NHANES IV and for the initial and final summary rows 1 and 6 for NHANES III and II. The unadjusted education gradients in prevalence in row 1 are large and generally statistically significant and have increased slightly over these 25 years. Even though the average age is approximately the same in all three education groups, controlling for age diminishes the schooling gradient in diabetes due to the nonlinear effect of age on prevalence with the least educated more likely to be the oldest. Approximately half of the schooling gradient with diabetes prevalence is accounted for by age, race, and ethnicity controls.

Smoking and exercise contribute to an additional diminishing of the education gradient, but this is largely offset by including past parental diabetes. Adding in the effects of being overweight or obese further attenuates the estimated gradient: combined, all of these behavioral and demographic controls explain somewhere between 60% and 75% of the schooling gradient, depending on which NHANES is examined in row 5. The estimated schooling gradient in row 5 appears now to increase only slightly over these 25 years.

When schooling and income are included as SES markers in the final row of Table 7, a small schooling gradient remains. Controls for income and education are different. Education has been shown to be related to the onset of diabetes but not income

Table 7. Contribution of factors toward measuring the schooling gradient in total diabetes prevalence

Other variables*	Coefficients on schooling					
	NHANES IV		NHANES III		NHANES II	
	Ed med	Ed high	Ed med	Ed high	Ed med	Ed high
(1) None	−0.058 (4.06)	−0.078 (6.04)	−0.045 (5.06)	−0.071 (2.81)	−0.019 (1.24)	−0.066 (4.08)
(2) = (1) + Race and ethnicity	−0.021 (1.55)	−0.041 (3.31)				
(3) = (2) + Smoking + exercise	−0.049 (1.38)	−0.036 (2.79)				
(4) = (3) + Parental diabetes	−0.025 (1.96)	−0.041 (3.43)				
(5) = (4) + Excessive weight	−0.020 (1.65)	−0.033 (2.89)	−0.012 (1.68)	−0.023 (2.95)	0.015 (1.02)	−0.014 (0.91)
(6) = (5) + Income groups	−0.015 (1.23)	−0.019 (1.62)				

Estimated DF/dx for schooling coefficients with z statistics based on robust standard errors in parentheses.

*Other variables included in model in addition to schooling.

(18). Instead, reduced income through lower ability to work appears to be a consequence of diabetes onset and not a cause. Based on that reasoning, income does not belong in these models. The education effects in row 5 represent my preferred summary of net effects of education on diabetes prevalence.

Conclusions

Although the increase in diabetes prevalence over time is considerably less than that indicated by the commonly used diagnosed diabetes, it remains a public health concern. The principal forces leading to higher diabetes prevalence are excessive weight and obesity and inheritance of diabetes through parents, which given the short time span studied most likely reflects a common environment or a gene environmental interaction. These forces were only partially offset by improvements in the education of the population over time.

Undiagnosed diabetes remains an important health problem with approximately one in five male diabetics undiagnosed in 1999–2002. This is far less of a problem than 25 years ago, when almost half of male diabetics were undiagnosed. Although race and ethnic differentials in undiagnosed diabetes were eliminated over the last 25 years, the disparities became larger across other measures of disadvantage such as education. Undiagnosed diabetes is a particularly severe problem among the obese, a group at much higher risk of diabetes onset.

Those in lower education groups face a triple diabetes threat. At least in more recent years, they are of slightly higher risk in

contracting the disease. Second, they remain at much greater risk of having their diabetes undiagnosed and presumably untreated. Third, even after diagnosis, they have considerable difficulty in successful disease management using the complex treatments necessary to diminish the negative health consequences associated with diabetes (19).

Partially counteracting these disturbing trends in diabetes prevalence, several recent studies have shown that health consequences of diabetes have declined over time. The relative mortality risks associated with obesity appear to have decreased significantly (15). In 2005, Gregg *et al.* (13) documented that CVD risk factors such as total cholesterol, blood pressure, and smoking (except diabetes) within BMI groups have declined, suggesting that the health consequences of obesity, although still severe, may also be declining. Other research has shown that there has been a significant reduction of incidence risk of CVD over time that has been at least as high among diabetics as for those who are not diabetic (3). These reductions in bad health consequences may be due to improvements in quality of care among diabetics (4) and better self management as education levels have improved (19).

I thank James Banks, Darius Lakdawalla, Raynard Kington, Meena Kumari, and David Weir for helpful comments. Iva MacLennan provided excellent programming assistance. This research was supported by grants from the National Institute on Aging.

1. US Department of Health and Human Services (2004) *Diabetes: A National Plan for Action: Steps to a Healthier US* (US Dept of Health and Hum Services, Washington, DC).
2. Banks J, Marmot M, Oldfield Z, Smith JP (2006) *J Am Med Assoc* 295:2037–2045.
3. Fox C, Coady S, Sorlie P, Levy D, Meigs J, D'Agostino R, Wilson P, Savage P (2004) *J Am Med Assoc* 292:2495–2499.
4. Saaddine J, Cadwell B, Gregg E, Engelgau M, Vinco G, Imperatore G, Narayan K (2006) *Ann Intern Med* 144:465–474.
5. Kanjilal S, Gregg E, Cheng Y, Zhang P, Nelson D, Mensah G, Beckels A (2006) *Arch Intern Med* 166:2348–2355.
6. US Department of Health and Human Services (1981) *Plan and Operation of the Second National Health and Nutrition Examination Survey 1976–1980* (US Dept of Health and Hum Services, Washington, DC), Programs and Collection Procedures Ser 1, No 15.
7. US Department of Health and Human Services (1994) *Vital and Health Statistics: Plan and Operation of the Third National Health and Nutritional Health and Nutrition Examination Survey, 1988–1994* (US Dept of Health and Hum Services, Washington, DC), Ser 1, No 32, DHHS No 94-1308.
8. Centers for Disease Control and Prevention and National Center for Health Statistics (2006) *National Health and Nutrition Examination Survey Data Sets and Related Documentation (Survey Questionnaire, Examination and Laboratory Protocols, 1988–1994 and 1999–2002)* (Natl Center for Health Stat, Hyattsville, MD).
9. National Center for Health Statistics (2004) *NHANES Analytical Guidelines* (Natl Center for Health Stat, Hyattsville, MD).
10. Imperatore G, Cadwell B, Geiss L, Saaddine J, Williams D, Ford E, Thompson T, Narayan K, Gregg E (2004) *Am J Epidemiol* 160:531–539.
11. Rashid I, Grossman M, Chou S (2005) *The Super Size of America: An Economic Estimation of Body Mass Index and Obesity in America* (Natl Bureau of Econ Res, Cambridge, MA), NBER Working Paper 11584.
12. Goldman N, Lin I, Weinstein M, Lin Y (2003) *J Clin Epidemiol* 56:148–154.
13. Gregg E, Cheng Y, Cadwell B, Imperatore G, Williams D, Flegal K, Narayan K, Venkat K, Williamson D (2005) *J Am Med Assoc* 293:1868–1874.
14. Gregg E, Cadwell B, Cheng Y, Cowie C, Williams D, Geiss L, Engelgau M, Vincer G (2004) *Diabetes Care* 27:2806–2812.
15. Flegal KM, Graubard BI, Williamson DF, Gail M (2005) *J Am Med Assoc* 293:1861–1867.
16. Kapteyn A, Smith JP, Van Soest A (2007) *Am Econ Rev* 97(1):461–473.
17. Lakdawalla D, Philipson T (2002) *The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination* (Natl Bureau of Econ Res, Cambridge, MA), NBER Working Paper 8946.
18. Smith JP (2004) *Popul Dev Rev Suppl: Aging, Health Public Policy* 30:108–132.
19. Goldman D, Smith JP (2002) *Proc Natl Acad Sci USA* 99:10929–10934.